

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ :

G06F 153/00, 17/30

A1

(11) International Publication Number:

WO 99/41694

(43) International Publication Date:

19 August 1999 (19.08.99)

(21) International Application Number: PCT/US99/03119

(22) International Filing Date: 12 February 1999 (12.02.99)

(30) Priority Data:

60/074,678

13 February 1998 (13.02.98)

US

(71) Applicant: YAHOO, INC. [US/US]; Suite 201, 3400 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventor: GRAHAM, Paul; 4175 Page Mill Road, Los Altos, CA 94022 (US).

(74) Agent: ALBERT, Philip, H.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).

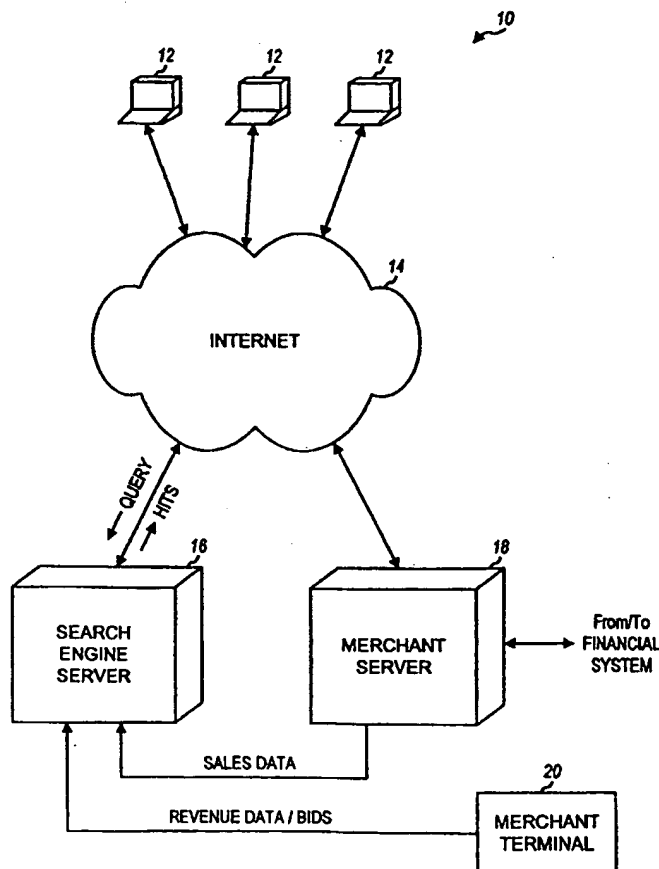
(81) Designated States: AU, CA, NZ, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published*With international search report.**Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: SEARCH ENGINE USING SALES AND REVENUE TO WEIGHT SEARCH RESULTS

(57) Abstract

A search engine (16) selects one or more search hits from among a plurality of hits, wherein a hit is a reference to a page or a site, based on a user interest, comprising an input module for accepting a query from a user, the query representing an interest of the user; a tracking module for tracking the user's navigation through the plurality of pages, including at least a destination purchase page, the destination purchase page being a page from which the user makes a purchase; a sales module which records associations between purchases and queries where the associations are provided, at least in part by an output of the tracking module; and a search module (16), which takes as its inputs at least a query and sales associations of that query provided by the sales module, and which outputs one or more search hits based on at least the query and the sales associations of that query. In some systems, instead of using sales data to alter the weights of the search results, merchant bidding is used to alter the weights of the search results, or a combination of the two is used.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

SEARCH ENGINE USING SALES AND REVENUE TO WEIGHT SEARCH RESULTS

5

BACKGROUND OF THE INVENTION

10 The present invention relates to search engines and more specifically to search engines used to locate goods and services on the Internet.

 A Web page is a sequence of text, optionally including images, that can be requested from a client, such as a Web browser, from a computer or Web server on a network. A Web site is a collection of stored or dynamically generated Web pages.

15 A URL is a string of characters that serves as the address of a Web page. A client sends a URL to a Web server, and receives the corresponding Web page in return.

 Some Web pages allow visitors to make purchases. For example, a Web page may contain information about an item for sale, along with a button that allows the
20 person seeing the page to place an order for it.

 A search engine is a program that helps users find information in a network of Web pages. Users submit to the search engine words or phrases indicating what they are searching for, and the search engine replies with a list of Web pages it predicts are relevant to that query. The page considered by a search engine for inclusion
25 in this list can be termed a "target page".

 The list of Web pages returned by a search engine is ranked by relevancy. Typically, relevancy is determined mostly by the content of the target pages.

 For example, if the user searches for "chocolate cake", a typical search engine will rank pages containing the phrase "chocolate cake" before those which merely
30 contain the words "chocolate" and "cake" separately, and those pages will in turn be ranked higher than pages that contain one of the two words but not the other.

One of the reasons people use search engines is to search for items for sale. A user who wants to buy a digital camera online will often begin by going to a search engine and using "digital camera" as the search phrase.

Existing search engines vary in how they rank pages. The rank of a target
5 page usually depends in part on a textual match with the search phrase. Search engines designed specifically for online shopping may also look at information on the target page, such as price and availability, and use that to determine the ranking. The Web site in which a page occurs may also influence the ranking of that page.

In general, a search occurs as follows: (1) the user submits a query Q to a
10 search engine, (2) the search engine returns a list of target pages, ranked based on their content, (3) the user goes to one of the pages in the list of search results and perhaps, while visiting the Web site containing that page, places an order. However, someone searching for the name of a product, like "digital camera", probably wants to buy it, and so wants to be given a list of pages where they can do so.

15 The search engines depend on site and page contents. For example, suppose that a search engine knows of two sites, A and B, that each have pages containing the phrase "digital camera". The two pages score the same for that phrase using whatever algorithm the search engine uses to rank pages. But whereas Site A is a well-designed site inspiring the consumer to buy, Site B is not well designed and does not
20 appeal to customers. To maximize sales, the search engine should preferably rank Site A before Site B. However, short of having a human look at both sites, it is difficult to ensure that the page from Site A appears ahead of the page from Site B in the search results.

SUMMARY OF THE INVENTION

25 A search engine selects one or more search hits from among a plurality of hits, wherein a hit is a reference to a page or a site, based on a user interest, comprising an input module for accepting a query from a user, the query representing an interest of the user; a tracking module for tracking the user's navigation through the plurality of pages, including at least a destination purchase page, the destination purchase page being a page
30 from which the user makes a purchase; a sales module which records associations between purchases and queries where the associations are provided, at least in part by an output of the tracking module; and a search module, which takes as its inputs at least a

query and sales associations of that query provided by the sales module, and which outputs one or more search hits based on at least the query and the sales associations of that query. In some systems, instead of using sales data to alter the weights of the search results, merchant bidding is used to alter the weights of the search results, or a

5 combination of the two is used.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

10 Fig. 1 is a block diagram of a computer network in which the present invention is used.

Fig. 2 is a logical diagram showing the interaction in one process using the elements of the apparatus shown in Fig. 1.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

15 In the figures, like numbers indicate like (but not necessarily identical) elements. In some cases, where many like elements are shown, not all of the like elements are called out with numbers, so that the figures are not unduly cluttered. In some instances, a number is followed by a letter in parenthesis to indicate a specific subclass of otherwise like elements.

20 With the present invention, users can search for pages related to goods and services by submitting a search phrase to a search engine and having the search engine return "hits" (relevant pages or sites) on that search phrase. The relevance of the hits presented are a function of the search phrase, but the relevance is adjusted, by weighting or otherwise, based on sales, revenue or bidding data. The sales data indicates

25 associations between search terms and purchases, such that when a search beginning with particular search term tends to result in a particular purchase, an association is noted in the sales data. That association is used to more heavily weight the pages or sites that correspond to purchases identified in the association. The revenue and bidding data is data provided by merchants used to modify weighting based on merchant interests.

30 One arrangement of computer elements implementing one aspect of the present invention is shown in Fig. 1. In a computer network 10, clients 12 are coupled

through Internet 14 to a search engine server 16 and a merchant server 18. In practice, many more clients would be connected than the three clients shown. Only one search engine server and one merchant server are shown, but more than one of each is possible.

Fig. 1 shows connections of each client 12 and servers 16, 18 to Internet 14 for intercommunication, as well as a connection between merchant server 18 and search engine server 16 for passing sales data to search engine server 16. Other connections shown include a connection between the merchant server and a financial system and a connection between a merchant terminal 20 and search engine server 16.

Several elements in the system shown in Fig. 1 are conventional, well-known elements that need not be explained in detail here. For example, a client 12 could be a desktop personal computer, workstation, cellular telephone, personal digital assistant (PDA), laptop, or any other computing device capable of interfacing directly or indirectly to the Internet. The present invention does not require the Internet, which refers to a specific global internetwork of networks, but is shown with the Internet as the mechanism for interconnection because that is the most likely mechanism where the present invention will be used. Notwithstanding the above, it should be understood that other networks can be used in place of Internet 14 in Fig. 1, such as an intranet, an extranet, a virtual private network (VPN), a non-TCP/IP based network, or the like. The interconnections between merchant systems and the search engine are shown outside Internet 14, but those connections might also be handled over Internet 14. Except for the details described herein and their equivalents, the invention can be implemented starting with a conventional merchant server with its connections to financial system, so further details of the precise operation of a merchant server need not be set out here. Client 12 typically runs a browsing program allowing a user of client 12 to browse pages available to it from servers connected to Internet 14.

While Fig. 1 illustrates one arrangement of hardware components, fig. 2 illustrates one process for passing data among those components according to one aspect of the present invention. The steps of the process are numbered in the order that they are likely to occur. The process illustrated in Fig. 2 begins with a user using a client to submit a query, Q, to a search engine (step 1). Typically, queries are in the form of search strings, such as "chocolate cake" or "digital video camera", but queries might take other forms, such as selections from pick lists or other data representing the interests of the user.

The search engine then uses query Q to come up with a hit list, H. Hit list H has entries that either represent specific pages, specific sites or both. Typically, a specific page is represented by a URL of the form "*http://hostdomain/path/filename*" and a site might be represented by a URL of the form "*http://hostdomain*". As explained below, hit list H is a list of those pages found to be weighted most relevant for query Q. Once hit list H is generated, it can be sent to the client for user review (step 2). In some cases, if the hit list is larger than a threshold number of hits, only part of the list is sent at a time. For example, if the list contains 500 hits, the initial set to the client might be only 20 hits, preferably the most relevant 20 hits.

When the client receives hit list H, in whole or part, it displays the received hits for user selection. The user selects a hit and the client navigates the user's browser to the page or site referenced by the hit by sending the URL for the hit to the appropriate server (step 3), typically a merchant's server if the user is looking for goods or services to purchase. Once the user navigates to the merchant server, the user will perform a purchasing interaction (step 4). Broadly construed, anything the user does at the merchant site is a purchasing interaction, such as looking around the site at the merchant's offerings, purchasing goods or services, or declining to make a purchase. The merchant server keeps track of the purchasing interactions and provides sales association data based on the purchasing interactions to the search engine (step 5). The merchant server can provide a record of what occurred for each interaction, or the merchant server can collect many records and send them in bulk to the search engine.

If the search engine and the merchant engine are operated by a single entity, that entity can identify which search terms go with which purchasing interactions by following the user. For example, a search/merchant engine could ask a user to log in for a session and each of the queries and hit lists could be recorded for a session. The engine would also track which purchases interactions were made with which pages selected from a hit list. In practice, however, the merchant servers will probably be independent of the search engine. To address this, URL encoding of the query might be used.

An example of URL encoding will be described, but many other methods could be used. With URL encoding, the hit list of URLs provided to the user point to particular hits (pages or sites) and contain additional information that either identifies the query itself or references a session in which the user made the search. For example, a

URL for a page on digital video cameras could include extra information representing the query "digital video camera" or just a query number that the search engine can use with a table lookup to identify the search term. Preferably, the queries are encoded in such a way that they cannot be easily forged by merchants seeking to skew searches to their sites.

The merchant server records the extra information that arrives with the URL that the user selected. When the user finishes the purchasing interaction, which the merchant hopes is a purchase, the result of the purchasing interaction is provided to the search engine server along with the extra information provided with the URL sent to the merchant server by the client.

As the process of Fig. 2 is repeated many times, over many users and many search terms, the search engine will approach a state where similarly minded purchasers will proceed through the process of Fig. 2 and find that they are being directed to hits that are highly relevant for what it is they are interested in purchasing. This is because the search engine returns a list of target pages, ranked by a combination of a textual match and additional data about the sales associated with the queries used by users.

One interesting side effect of the process described above is that uninteresting sites will be filtered out, if few people are interested in the offerings at those sites. Also, common spelling errors will tend to get fixed. For example, if users searching for "compact disc player" end up spending considerable money at sites offering compact disc players, then those pages will have a higher relevance for that search phrase, even though the phrase "compact disc player" is not present on those pages.

In one embodiment of the invention, the data sent back from the target site to the search engine includes, for each search phrase, the number of visitors arriving with that search phrase encoded in the URL and for each item ordered by a visitor, the search phrase used, the amount spent, the URL of the arrival page, the URL of the page where the item was ordered. This information is used to influence the rankings of target pages as follows: In addition to whatever score a target page might have received for Q from the search engine's content-based ranking method, a target page gets another score consisting of an estimate of the amount of revenue that will be generated, on average, if the user goes to the target page. In one embodiment, these two scores are combined to produce the score used for ranking the target page.

In one embodiment, the method for estimating the revenue that will be generated from a target page is based on the average sales per capita from orders placed from that target page, by visitors whose queries match Q, during some past time period. Another variant would be to calculate sales per capita from users who arrived through that target page.

The rankings produced by this method differ from rankings based mainly on textual matches. In one possible example, if it turned out that users searching for "chocolate" spent the most money on pages selling diet pills, then such pages are given a high rank, even if they did not contain the phrase "chocolate" at all.

Data about what users do when they arrive at target sites is fed back to the search engine and used to adjust the rankings of the target sites. A site that would do well based on textual match, but which ultimately, for whatever reason, does not give the users what they were looking for, will tend to slip down in the rankings. Sites that consumers do buy from will tend to rise.

This provides a benefit for the search engines, in that (with no human intervention) the system naturally tends toward an equilibrium that maximizes their revenues. It is also a benefit for the consumer, since for consumers doing product searches, the ranking of sites is based on whether they found what they wanted.

In one embodiment of the invention, because target sites may offer search engines varying percentages of their sales, the number used by the search engine to rank a page should be the average sales per capita from past users searching for that (or related) phrases, multiplied by the percentage of sales offered by the target site.

The search engine can operate at three different levels of precision. At a first level, the search engine uses the sales association data fed back from the merchant server to rank individual pages in a target site associated with query Q. At a second level, the search engine can use the sales association data to calculate an overall average revenue per capita for query Q at the target site and use that to adjust the relevance of ranking of any page in that site. At the next level, the search engine uses the sales association data to calculate an overall average revenue per capita at the target site for all queries, and uses that value to influence the ranking of any page in that site. The appropriate level of precision depends on sample size. These calculations would be useful even in the last, most general case (Level 3) in that it would tend to eliminate the sites that did not impress visitors.

In one embodiment, data about recent orders is weighted more heavily than older data. In another embodiment, some amount of randomness is included in the ranking of target pages, in order to prevent low ranking pages from becoming too lowly ranked just because of their initial ranking. One way to include randomness would be, for every N-th search, to randomly reshuffle the presentation order of the top-ranking M target pages, for some N and M.

The methods described above provide a way for new sites to get traffic despite having no sales initially. One possible approach is to include some pages with zero sales in the random reshuffling described above.

In a further embodiment, the method is refined by defining the denominator for sales per capita as the number of times a link to the target page was shown to users at the search engine (perhaps weighted by placement on the page), rather than the number of times users chose to visit that target page. Though more costly to calculate, this approach comes closer to the ideal ranking of target pages to maximize revenue.

In calculating the ranking of a target page P for a query Q, the weighting of the sales-based component should increase with the sample size (the number of orders placed by users who searched for queries matching Q at the site containing P) and the frequency of queries matching Q. If many users search for Q, and many of those in turn buy something at P, then P ranks high in a search for Q.

In the case of Web sites with dynamically generated URLs that contain more information than necessary to identify the page, the URLs for pages kept at the search engine and sent back in the data should be base URLs that uniquely identify each page.

Another beneficial side effect of the above-described methods is that a search engine can be used for noncommercial as well as commercial searches, using the same criteria. For example, users searching for "election results" will spend little money, if any, as a result of their searches. The resulting revenue per capita for that key phrase will be low over all hits, and any revenue-based score would not contribute significantly to the ranking of the pages for that query.

In addition to using purchasing interaction information to alter the relevance weighting of hits, merchants might be allowed to alter the relevance weightings

by bidding for search terms. In the system of Fig. 1, this information might be supplied via merchant terminal 20.

The above description is illustrative and not restrictive. Many variations of the invention will become apparent to those of skill in the art upon review of this disclosure. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

WHAT IS CLAIMED IS:

1 1. A search engine for selecting one or more search hits from among a
2 plurality of hits, wherein a hit is a reference to a page or a site, based on a user interest,
3 comprising:
4 an input module for accepting a query from a user, the query representing an interest
5 of the user;
6 a tracking module for tracking the user's navigation through the plurality of pages,
7 including at least a destination purchase page, the destination purchase page
8 being a page from which the user makes a purchase;
9 a sales module which records associations between purchases and queries where the
10 associations are provided, at least in part by an output of the tracking module;
11 and
12 a search module, which takes as its inputs at least a query and sales associations of
13 that query provided by the sales module, and which outputs one or more search
14 hits based on at least the query and the sales associations of that query.

15 2. The search engine of claim 1, wherein the query is a search text string.

16 3. A search engine for selecting one or more search hits from among a
17 plurality of hits, wherein a hit is a reference to a page or a site, based on a user interest,
18 comprising:
19 an input module for accepting a query from a user, the query representing an interest
20 of the user;
21 a bidding module for tracking bid for associations between query phrases and
22 references; and
23 a search module, which takes as its inputs at least a query and associations of that
24 query provided by the bidding module, and which outputs one or more search
25 hits based on at least the query and the associations of that query.

1/2

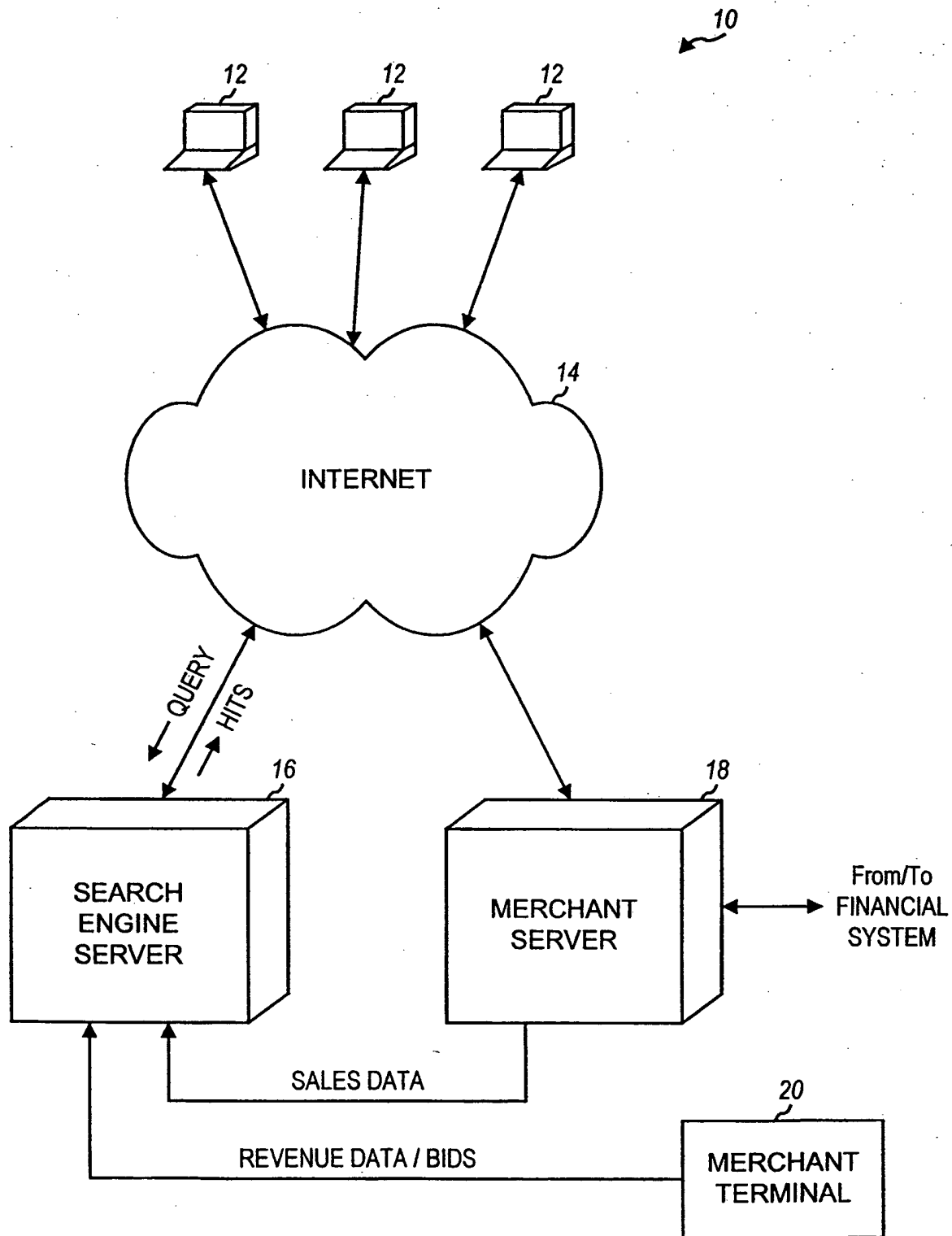
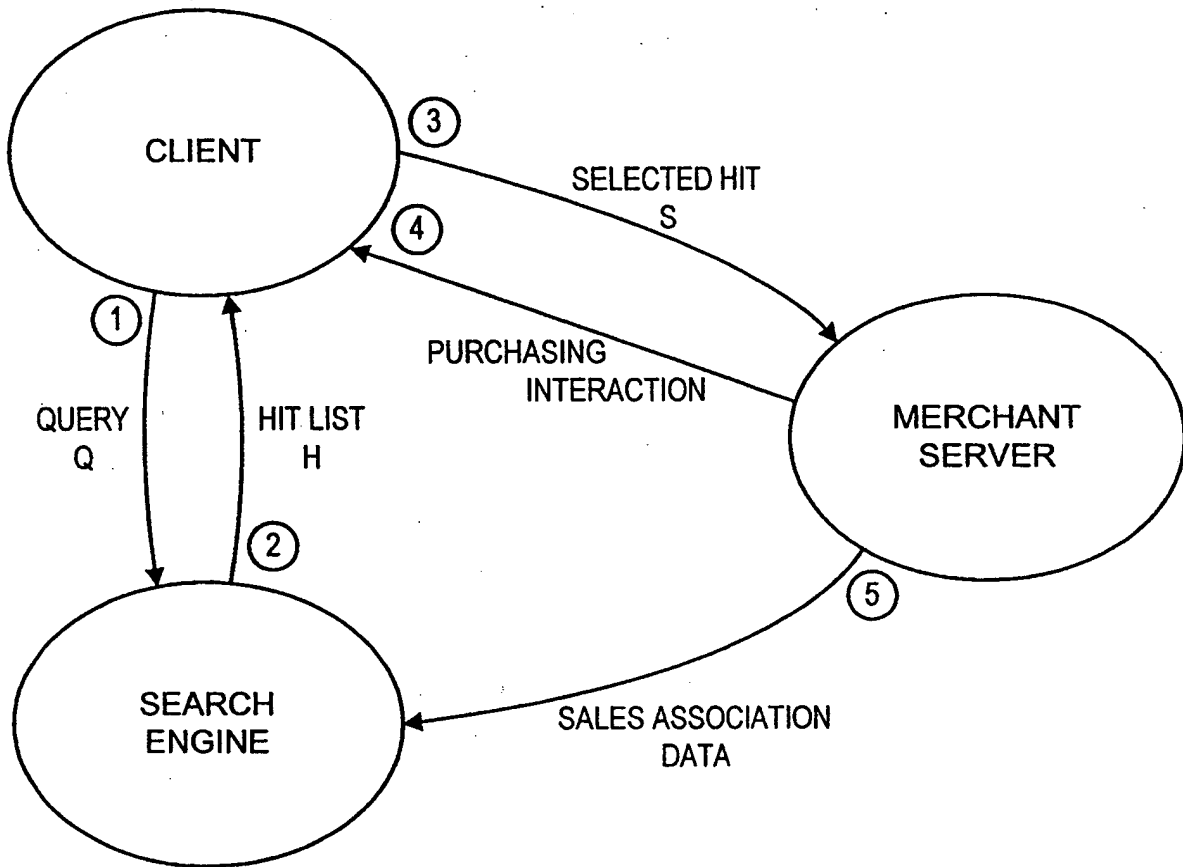


FIG. 1

2/2

**FIG. 2**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/03119

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : GO6F 153/00; 17/30

US CL : 705/26, 27; 380/23, 24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 705/26, 27; 380/23, 24

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Internet, DIALOG

search terms: input, search engines, text, sales, records, purchases, transactions, revenue, bidding, weight, navigate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, E --- Y	US 5,895,454 A(HARRINGTON) 20 April 1999, abstract, lines 7-10; col. 1, lines 40-54; col. 4, lines 11-15; col. 7, lines 2-8.	1, 2 ----- 3
Y, P	US 5,864,822 A (BAKER, III) 1 January 1999, abstract, lines 5-14.	3
A	M2 Presswire, BUY IT ONLINE: Buy It onLine adds new search engine, October 30, 1996 pages 1-3.	1-3

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

20 JUNE 1999

Date of mailing of the international search report

02 JUL 1999

 Name and mailing address of the ISA/US
 Commissioner of Patents and Trademarks
 Box PCT
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JAMES P. TRAMMELL

Joni Hill

Telephone No. (703) 305-3900